Information Seeking as Chasing Anticipated Prediction Errors

Jian-Qiao Zhu (j.zhu@warwick.ac.uk)
Wendi Xiang (wendi.xiang11@gmail.com)
Elliot A. Ludvig (e.ludvig@warwick.ac.uk)
Department of Psychology, University of Warwick
Coventry, UK CV4 7AL

Abstract

When faced with delayed, uncertain rewards, humans and other animals usually prefer to know the eventual outcomes in advance. This preference for cues providing advance information can lead to seemingly suboptimal choices, where less reward is preferred over more reward. Here, we introduce a reinforcement-learning model of this behavior, the anticipated prediction error (APE) model, based on the idea that prediction errors themselves can be rewarding. As a result, animals will sometimes pick options that yield large prediction errors, even when the expected rewards are smaller. We compare the APE model against an alternative information-bonus model, where information itself is viewed as rewarding. These models are evaluated against a newly collected dataset with human participants. The APE model fits the data as well or better than the other models, with fewer free parameters, thus providing a more robust and parsimonious account of the suboptimal choices. These results suggest that anticipated prediction errors can be an important signal underpinning decision making.

Keywords: information seeking; early resolution of uncertainty; anticipated prediction errors; forward sampling.

Introduction

Humans and other animals have a strong preference for informative options. They are inherently curious and will explore unknown options, even sacrificing rewards to resolve an uncertain outcome early. Sometimes the search for predictive information can be independent of profit and have no effect on the delivery of primary rewards, as if consuming information itself was rewarding (Wyckoff, 1952; Prokasy, 1956; Bromberg-Martin & Hikosaka, 2009; Iigaya, Story, Kurth-Nelson, Dolan, & Dayan, 2016). On occasion, this information seeking can lead to seemingly suboptimal behaviors with animals preferring options with lower expected values (Spetch, Belke, Barnet, Dunn, & Pierce, 1990; Roper & Zentall, 1999). In this paper, we develop a new computational model of this information-seeking behaviour based on the idea that animals' choices reflect both the expected rewards and the anticipated prediction errors from any upcoming cues.

This preference for advanced information has been widely observed across species, including rats (Prokasy, 1956; Chow, Smith, Wilson, Zentall, & Beckmann, 2016), pigeons (Spetch et al., 1990), starlings (Vasconcelos, Monteiro, & Kacelnik, 2015), monkeys (Bromberg-Martin & Hikosaka, 2009, 2011; Blanchard, Hayden, & Bromberg-Martin, 2015), and humans (Iigaya et al., 2016). In some cases, animals even give up food or water for advance information about impending rewards, even though these advanced signals do not change the eventual reward. For example, pigeons reliably choose an alternative that provides delayed access to food 50% of the time over one that always provides the same amount of food with the same delay, but only when an immediate cue is provided, which signals to the pigeons whether or not food will eventually be available on that trial (Spetch et al., 1990; Gipson, Alessandri, Miller, & Zentall, 2009). The choice of the 50% option is clearly suboptimal in terms of reward-intake maximization. Similarly, when choosing between delayed, probabilistic rewards, monkeys and humans will prefer an option that informs them about the eventual outcome of that trial over one that leaves the resolution of uncertainty to the time of reward delivery (Bromberg-Martin & Hikosaka, 2009, 2011; Iigaya et al., 2016).

In addition to the presence of advance information, a few variables have proven critical to the emergence of this suboptimal choice (see McDevitt, Dunn, Spetch, and Ludvig (2016) for review). First, the contingencies between the predictive cues and the outcomes is important because it influences the amount of uncertainty resolved by the cues: The more information conveyed by the predictive cues, the more preferred the associated choice target (Bromberg-Martin & Hikosaka, 2009). Second, humans and other animals also exhibit a preference for earlier advanced notice of the eventual outcome. Increased information seeking has been found in the case of longer delays (Spetch et al., 1990; Iigaya et al., 2016). Third, the subjective value of advance information scales with the reward magnitude of the potential outcomes (Blanchard et al., 2015). Finally, aversive outcomes can sometimes produce outright information avoidance in human subjects, as in the *Ostrich effect* (Karlsson, Loewenstein, & Seppi, 2009).

Given this rich set of empirical findings, we endeavored to build a computational model that can capture as many of these empirical results as possible, but first we briefly review the existing computational models for this information-seeking behaviour.

Existing Computational Models

The apparent departures from optimality observed when advance information is available poses a significant computational challenge to standard models of reinforcement learning (RL) (Niv & Chan, 2011). Previous research has explored several possible extensions and refinements to the usual RL framework, including the information bonus model (Bromberg-Martin & Hikosaka, 2011), the disengagement model (Beierholm & Dayan, 2010), and the anticipatory utility model (Iigaya et al., 2016). The information bonus model encapsulates the idea that receiving advance information acts as if it were a primary reward. This information bonus has alternatively been operationalized as either a free parameter

(Bromberg-Martin & Hikosaka, 2011) or as the Shannon entropy of the reward probability (Bennett, Bode, Brydevall, Warren, & Murawski, 2016). These ideas successfully explain the observed preference for more informative options (Bromberg-Martin & Hikosaka, 2009). Formalizing the information bonus as the Shannon entropy, however, fails to deal with, for instance, the fact that animals prefer to observe more even when the number of bits they receive by doing so is less (Roper & Zentall, 1999). On the other hand, without using Shannon entropy, the information bonus cannot capture the relationship between information seeking and probability, which resembles an inverse U-shaped function (Green & Rachlin, 1977).

The anticipatory utility model is a recently proposed alternative model for these data, which formalizes the economic idea of savouring (Loewenstein, 1987; Iigaya et al., 2016). According to this model, animals are hypothesized to enjoy or savour the anticipation of guaranteed rewards to come. Anticipatory utility alone, however, cannot explain why the delay to reward would influence how much the informative option is preferred (Spetch et al., 1990; Iigaya et al., 2016). This limitation emerges because delay renders anticipatory utility less rewarding at the same speed as the primary reward. To rectify this, an additional boosting mechanism was introduced to enhance anticipatory utility, and thereby slow down effect of discounting future rewards (Iigaya et al., 2016). The full model, including this boosting mechanism, explains a wide range of information-seeking behaviours, including many of the properties of sub-optimal choice. One challenge for the anticipatory utility model is how such a mechanism could be learned locally, as the computations require full knowledge of the eventual time to reward in advance (Niv & Chan, 2011).

The Anticipated Prediction Error Model

Given these limitations on prior models, here, we develop an alternative formalism centered around the idea of anticipated prediction errors (APE). According to the APE model, animals draw one-step samples of their anticipated futures from a simple model of the world and calculate the prediction error that would be associated with that sample. These anticipated prediction errors are then treated as though they were rewarding in and of themselves, reminiscent of how momentary subjective well-being correlates with prediction errors (Rutledge, Skandali, Dayan, & Dolan, 2014). These samples are biased such that futures which contain positive prediction errors are more likely to be sampled. This *forward sampling* (i.e. anticipation) from the current state using imagined experiences and learned environmental dynamics, such as developed in the Dyna-2 architecture (Silver, Sutton, & Müller, 2008), can provide useful anticipatory signals that guide decision making. The critical difference between the APE model and the standard RL model is that the APE model maintains two separate valuation systems: one estimated from actual experience (model-free), and the other estimated through this forward-sampling process (model-based). The prediction errors generated via the forward traces are called anticipated prediction errors (APEs). Together with the conventional value functions, these APEs drive the preference to seek or avoid certain future states. The bias in the sampling process toward positive prediction errors can even induce suboptimal choices.

Model Specification

We extend the standard Temporal-difference (TD) model (Sutton & Barto, 1998) where agents are assumed to estimate an action-value function for each experimental stimulus:

$$Q(s_t, a_t) = \mathbb{E}[\sum_{k=1}^{\infty} \gamma^k r_{t+k-1}] \quad (1)$$

where t indexes time, s_t specifies the state visited at time t, r_t indicates the immediate reward delivered at time t, and $\gamma \in [0,1)$ is a discount factor, which devalues delayed rewards. This action-value function represents the expected discounted future reward. In TD learning, this action-value function is estimated through a simple incremental update mechanism:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\delta_{t+1} \quad (2)$$

where α is the learning rate and δ is the reward prediction error (RPE), calculated as follows:

$$\delta_{t+1} = r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \quad (3)$$

This RPE signal represents the difference between the value of the current state-action pair and the value of the best next-state-action plus the reward achieved in the transition. Thus, RPEs are triggered by each state transition; the mechanics of the APE model hinge on the transition from choice to the predictive cues which reveals what the eventual outcomes will be on that trial. In particular, a good cue, which resolves reward uncertainty appealingly, will generate positive RPEs, whereas a bad cue will generate negative RPEs. The RPEs will be zero in response to non-predictive cues, once the values are well learned.

Here, we define anticipated prediction errors (APE) as the perceived discrepancy between the current state (what it is like at present) and an anticipated future state (how it would be in the future) (see Figure 1). Formally, if there is no immediate primary reward delivered during the trajectory from state s to s' (e.g., the transition from choice state to cue states in the information choice task), then the value of APE in state s when anticipating future state s' is defined as the product of prediction errors between the two states and the transition probability:

$$APE(s'|s,a) = T(s'|s,a) \times [\gamma^{D_{ss'}} \max_{a'} Q(s', a') - Q(s,a)] \quad (4)$$

where $D_{ss'}$ is the time taken to travel from s to s', and $T(s'|s,a)$ is the transition probability from s to s' by taking action a. In the simulations here, this travel time is always taken to be 1, but the formulation is more general.

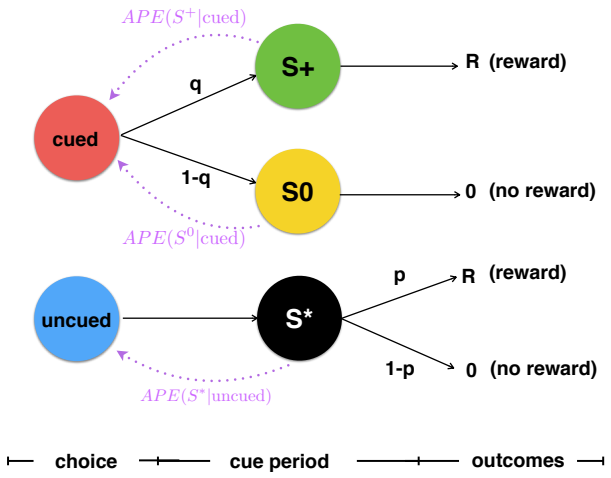

Figure 1: Formal representation of the information-choice task as a Markov Decision Process (MDP). Two offers (red and blue circles) are presented and the animal must choose one of them. A cue then appears after this initial choice, which is either informative (green S^+ indicates a rewarding outcome, and yellow S^0 indicates a neutral outcome) or uninformative (black S^* leaving the animal in a state of uncertainty). Following a delay (T_{delay}), the animal obtains the outcome (reward or no reward). The anticipatory signals proposed by the APE model are illustrated as purple dashed lines.

Note that this computation relies on the samples generated based on the learned environment dynamics. The primary assumption of the APE model is that humans and other animals treat APEs as though they were rewarding, whereby positive APEs are reinforcing and negative APEs punishing. The APEs are positive when the anticipated value of the future sampled state is better than value of the present state and negative in the opposite case. These quantities can also be understood as measurements of the pleasure (displeasure) one derives from anticipating the good cue (bad cue). Furthermore, attention weights are assigned to each individual APE, specifying the relative likelihood that a particular future state will be sampled. Accordingly, the decision value $\bar{Q}$ of taking action a is defined as the weighted sum of APEs for anticipated future outcomes plus the value function for the corresponding state:

$$\bar{Q}(s,a) = \sum_{s_k \in \mathcal{S}} w_k APE(s_k|s,a) + Q(s,a) \tag{5}$$

where $\mathcal{S}$ denotes the set of all possible future states after taking action a at the state s that a subject can attend to.

Given the decision values of both the cued and the uncued options, the softmax function is then used to compute the probability of choosing the cued option:

$$P(a) = \frac{e^{\beta \bar{Q}(s,a)}}{\sum_{a' \in \mathcal{A}} e^{\beta \bar{Q}(s,a')}} \tag{6}$$

where $\mathcal{A}$ is the set of all possible actions at state s, and β is an inverse temperature parameter, which controls the degree of exploration.

Experiment

We conducted an empirical experiment to evaluate the quality of the APE model in comparison with the information-bonus model discussed above. In the experiment, people were repeatedly given a choice between an informative or non-informative option, where the outcomes were delayed 20s. Outcomes were either positive (erotic images), neutral (images of objects) or negative (aversive images). Good trials involved positive or neutral images, Mixed trials involved positive or negative images, and bad trials involved negative or neutral images. These outcomes were always delivered with 50/50 odds on each trial. Qualitatively, the APE model predicts that people will seek information in the positive and mixed cases, but not the negative cases. This prediction emerges from bias toward sampling future states with positive outcomes. The information bonus model would expect equivalent information seeking in all cases, as the amount of information present is equal in all three types of choices.

Methods

Participants Eighty human participants were recruited from the Warwick University SONA system. All participants gave informed consent and were paid a flat rate of 5 pounds for their participation.

Task Participants performed the experiment on Windows PCs running PsychoPy (Peirce, 2007). The task was a simple two-alternative forced choice between an uncued target (*Keep It Secret*), which was followed by a non-predictive cue, and a cued target (*Find Out Now*), which was immediately followed by predictive cues that signalled the eventual outcome. Choosing either the cued or the uncued option did not alter the odds nor the eventual outcomes. The only difference between the two options was the presence or absence of advance information about those eventual outcomes. After choice, the cue was present for 20 seconds in all trials. The outcome image was presented immediately at the end of this cue. To ensure participants viewed the image, they had to press a randomly selected key (indicated on the image proper) to advance to the next trial.

The experiment consisted of three different conditions in terms of the valence of eventual outcomes. In the Good condition, the gamble included 50% erotic images and 50% neutral images (as illustrated in Figure 2). In the Bad condition, the gamble included 50% aversive images and 50% neutral images. In the Mixed condition, the gamble included 50% erotic images and 50% aversive images. The images used in the experiment were previously validated in the International Affective Picture System (Lang, Bradley, Cuthbert, et al., 1999). Sixteen images from the "EroticCouple" category were selected as positive images for heterosexual subjects and another 16 images from "Mutilation" category were selected as the aversive images. Images were chosen as the rewards

so that they could be consumed immediately, as opposed to monetary rewards (Crockett et al., 2013). All participants completed 16 interleaved trials for each condition, making 48 trials in total. Participants were informed about the nature of the potential outcomes before the experiment started.

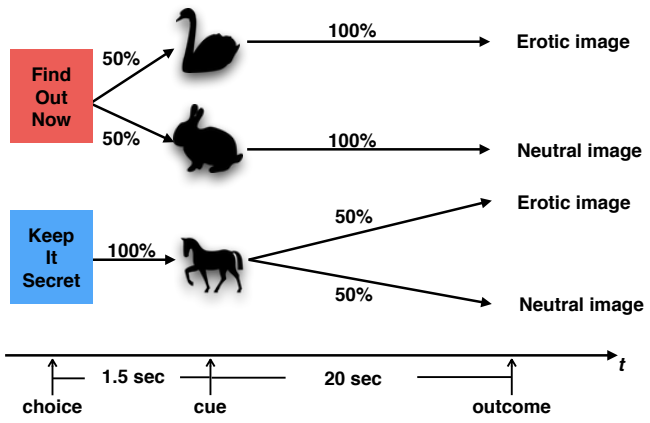

Figure 2: Human information choice task. The diagram illustrates the Good condition, which contains a gamble of 50% erotic and 50% neutral images. The experiment also tested a Bad condition (50% aversive and 50% neutral images) and a Mixed condition (50% erotic and 50% aversive images).

Results

A total of 69 heterosexual participants (48 female and 21 male) completed the task. Eleven participants were excluded (6 non-heterosexual, 4 did not disclose their sexual orientation, and 1 did not complete the task). Only the data from the last nine trials per condition are reported here.

As shown in Figure 3, participants chose the cued option on average $42.8\% \pm 8.2\%$, $60.2\% \pm 7.6\%$, and $71.8\% \pm 6.5\%$ in the Bad, Mixed, and Good conditions respectively. Choices in the Good condition and the Mixed condition were significantly higher than chance responding (Good: $t(68) = 6.72, p < 0.001, d = 1.143$; Mixed: $t(68) = 2.68, p = 0.009, d = 0.456$). In the Bad condition, people chose the informative slightly below chance, but not significantly so ($t(68) = -1.75, p = 0.085, d = -0.297$). There were, however, considerable individual differences in the preferences for advance information (dashed grey lines in Figure 3). This pattern of responses qualitatively agree with the predictions of the APE model, but not the information-bonus model.

Model Comparisons

Next, we attempted a quantitative model comparison, fitting both the APE model and the information-bonus model to the individual choice proportions in the current dataset.

To fit the APE model to the data, first note that the expected rewards for both options are held constant in the experimental task, $Q(\text{cued}) = Q(\text{uncued})$. In addition, receiving non-predictive cues leaves participants equally uncertain about the eventual outcomes, and thus sampling from those states does not generate any anticipated prediction errors, $APE(S^*|\text{uncued}) = 0$. This analysis suggests that only the APEs related to the predictive cues determine choices in the current task. Following this logic, from equation (5), we can calculate differences in the action values of the cued and uncued options in the Good, Bad, and Mixed conditions, respectively as follows:

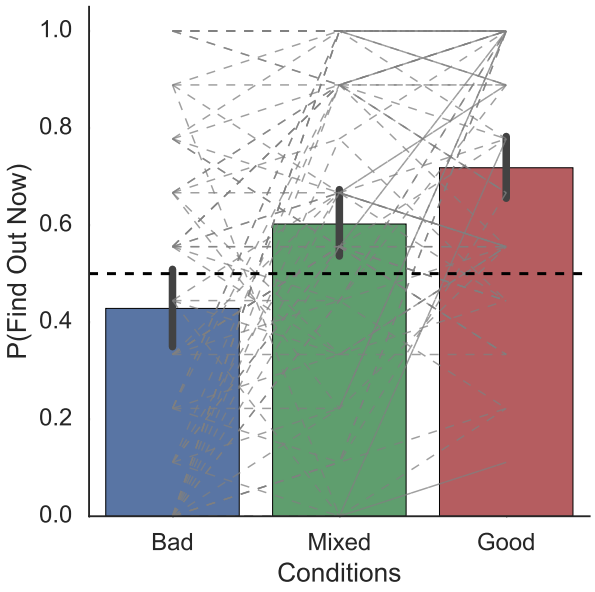

Figure 3: Mean percentage of choosing cued option (*Find Out Now*) in the Bad, Mixed, and Good conditions. Error bars indicate $\pm$ SEM in mean choice proportions. The dashed black line indicates the 50% choice probability. The dashed grey lines are individual choice probabilities.

$$\Delta\bar{Q}_{\text{Good}} = w^+ APE(S^+|\text{cued}) + w^0 APE(S^0|\text{cued}) \quad (7)$$

$$= (w^+ - w^0)\frac{\gamma^T R}{4} \quad (8)$$

$$\Delta\bar{Q}_{\text{Bad}} = w^- APE(S^-|\text{cued}) + w^0 APE(S^0|\text{cued}) \quad (9)$$

$$= (w^0 - w^-)\frac{\gamma^T R}{4} \quad (10)$$

$$\Delta\bar{Q}_{\text{Mixed}} = w^+ APE(S^+|\text{cued}) + w^- APE(S^-|\text{cued}) \quad (11)$$

$$= (w^+ - w^-)\frac{\gamma^T R}{2} \quad (12)$$

where T is the length of the delay, R is the absolute magnitude of rewards or punishment, and S^+, S^-, S^0 are the cues indicating positive, negative, or neutral outcomes. The weight factors are associated with their corresponding future states. Note that only the differences in weights matter for model behavior.

Any individual differences are reflected by the weight pa-

rameters in the APE model. The APE model predicts no preference for advance information when $w^+ = w^- = w^0$. We hypothesize that the differences in weights for various future outcomes give rise to information seeking or avoidance behaviors. The preferences for advance information would arise, for instance, in the Mixed condition if the model weights S^+ more heavily than S^-: $w^+ > w^-$.

As described above, the information-bonus model (Bromberg-Martin & Hikosaka, 2011) assumes that information has an intrinsic value, r_{info}, which in our setting was delivered upon each state transition to an informative cue state. This model would predict the differences in decision values as follows:

$$\Delta\tilde{Q} = r_{\text{info}} \tag{13}$$

for all situations where information is sometimes available. We also considered a potential extension to the information-bonus model which would assign different values to different types of information: $r^+_{\text{info}}, r^-_{\text{info}}, r^0_{\text{info}}$ for viewing reward predicting cues, punishment predicting cues, and neutrality predicting cues respectively.

Model Fitting

The models were fit to the the data using hierarchical Bayesian modeling (Huys et al., 2011), in which the maximum a posteriori estimate of each parameter h_i for each participant i is calculated. These parameters are treated as a random sample from a Gaussian population distribution with means and variance $\theta = \{\mu_\theta, \Sigma_\theta\}$. Model comparison was based on the integrated Bayesian Information Criteria (iBIC) scores with an uninformative prior. As such, we analyzed the log likelihood $p(D|M)$ of each model directly:

$$p(D|M) = \int p(D|\theta)p(\theta|M)d\theta \tag{14}$$

$$\approx -\frac{1}{2}\text{iBIC} \tag{15}$$

$$= \log p(D|\theta^{ML}) - \frac{1}{2}|M|\log|D| \tag{16}$$

where $|D|$ is the number of choices made by all participants, and $|M|$ is the number of parameters fitted. We compute the $\log p(D|\theta^{ML})$ as the sum of integrals over individual parameters:

$$p(D|\theta^{ML}) = \sum_i \log \int p(D_i|h)p(h|\theta^{ML})dh \tag{17}$$

$$\approx \sum_i \log \frac{1}{K}\sum_{k=1}^{K} p(D_i|h_k) \tag{18}$$

where the integrals are replaced by a sum over samples from the empirical prior. This step ensures that we evaluate how well the model fits the data only using information about the group parameters.

As a result, the iBIC penalizes for model complexity, and the model with the lowest iBIC is taken as the best-fitting model. As shown in Table 1, the best-fitting APE model vastly outperformed the different information-bonus models. We use the iBIC score of the best fitted model as a baseline and derive the differences in iBIC as ΔiBIC.

Table 1: Quality of model fit to behavioral data

Model	Free parameters	Model iBIC	ΔiBIC
APE	$w^+ - w^-, w^- - w^0$	2065.1	40.4
APE	$w^+ - w^-, w^- - w^0, \gamma$	2126.3	101.6
APE	$w^+ - w^-, w^- - w^0, \beta$	**2024.7**	**0.0**
APE	$w^+ - w^-, w^- - w^0, \beta, \gamma$	2137.2	112.5
Info Bonus	r_{info}	2361.5	336.8
Info Bonus	r_{info}, β	2430.8	406.1
Info Bonus	$r^+_{\text{info}}, r^-_{\text{info}}, r^0_{\text{info}}$	2098.8	74.1
Info Bonus	$r^+_{\text{info}}, r^-_{\text{info}}, r^0_{\text{info}}, \beta$	2081.0	56.3

Discussion

We have introduced a novel model of information-seeking in choice, which assumes that preferences are driven by anticipated prediction errors (APEs) accumulated through simulated forward trajectories. These APEs are treated like rewards, which combined with a bias toward sampling trajectories with positive outcomes, leads to information seeking in situations with potential positive outcomes. The model was compared against an information-bonus model through a novel empirical experiment, whereby people were given the opportunity to get early information about rewarding or aversive outcomes. As the APE model predicted, and contrary to the information-bonus model, people only sought early information for positive outcomes. Quantitative model selection supported these conclusions.

In addition to better fitting the novel dataset, the APE model provides potential insights into other types of information-induced sub-optimal choices (McDevitt et al., 2016). For example, the positive APE scales with the probability of reward (larger with lower probabilities), providing a mechanism through which a lower probability reward could be preferred to a higher probability one, as sometimes observed in animals (Spetch et al., 1990; Roper & Zentall, 1999; Gipson et al., 2009). In addition, unlike an information bonus, the APE is sensitive to the magnitude of reward and would grow with larger rewards leading to greater preference for informative options, as observed with information-seeking in monkeys (Blanchard et al., 2015). Future work will require direct simulation of these other findings, as well as further comparison to existing models, including the different information-bonus models (Bromberg-Martin & Hikosaka, 2011; Bennett et al., 2016) and the anticipatory utility model (Iigaya et al., 2016).

The current experimental protocol only involved a shallow decision tree, and the corresponding APE model presented here only used one-step anticipation. For decision

trees with high branching factors and/or larger depths, however, it would be computationally intractable to sample from all possible forward trajectories. For example, one recent study used a four-stage information-seeking game, and observed systematic deviations from the optimal strategy (Hunt, Rutledge, Malalasekera, Kennerley, & Dolan, 2016). This type of task poses yet another computational challenge for all the models discussed here. The APE model, which already involves look-ahead experiences, is readily adaptable to incorporate other, more sophisticated, planning algorithms such as Monte-Carlo tree search (Coulom, 2006). This potential extension of the model to more complex tree search remains a question for further research.

References

Beierholm, U. R., & Dayan, P. (2010). Pavlovian-instrumental interaction in observing behavior. *PLoS Comput Biol*, *6*(9), e1000903.

Bennett, D., Bode, S., Brydevall, M., Warren, H., & Murawski, C. (2016). Intrinsic valuation of information in decision making under uncertainty. *PLoS Comput Biol*, *12*(7), e1005020.

Blanchard, T. C., Hayden, B. Y., & Bromberg-Martin, E. S. (2015). Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron*, *85*(3), 602–614.

Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, *63*(1), 119–126.

Bromberg-Martin, E. S., & Hikosaka, O. (2011). Lateral habenula neurons signal errors in the prediction of reward information. *Nature neuroscience*, *14*(9), 1209–1216.

Chow, J. J., Smith, A. P., Wilson, A. G., Zentall, T. R., & Beckmann, J. S. (2016). Suboptimal choice in rats: incentive salience attribution promotes maladaptive decision-making. *Behavioural Brain Research*.

Coulom, R. (2006). Efficient selectivity and backup operators in monte-carlo tree search. In *International conference on computers and games* (pp. 72–83).

Crockett, M. J., Braams, B. R., Clark, L., Tobler, P. N., Robbins, T. W., & Kalenscher, T. (2013). Restricting temptations: neural mechanisms of precommitment. *Neuron*, *79*(2), 391–401.

Gipson, C. D., Alessandri, J. J., Miller, H. C., & Zentall, T. R. (2009). Preference for 50% reinforcement over 75% reinforcement by pigeons. *Learning & Behavior*, *37*(4), 289–298.

Green, L., & Rachlin, H. (1977). Pigeons'preferences for stimulus information: Effects of amount of information1. *Journal of the experimental analysis of behavior*, *27*(2), 255–263.

Hunt, L. T., Rutledge, R. B., Malalasekera, N., Kennerley, S. W., & Dolan, R. J. (2016). Approach-induced biases in human information sampling. *bioRxiv*, 047787.

Huys, Q. J., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Comput Biol*, *7*(4), e1002028.

Iigaya, K., Story, G. W., Kurth-Nelson, Z., Dolan, R. J., & Dayan, P. (2016). The modulation of savouring by prediction error and its effects on choice. *eLife*, *5*, e13747.

Karlsson, N., Loewenstein, G., & Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and uncertainty*, *38*(2), 95–115.

Lang, P. J., Bradley, M. M., Cuthbert, B. N., et al. (1999). International affective picture system (iaps): Instruction manual and affective ratings. *The center for research in psychophysiology, University of Florida*.

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal*, *97*(387), 666–684.

McDevitt, M. A., Dunn, R. M., Spetch, M. L., & Ludvig, E. A. (2016). When good news leads to bad choices. *Journal of the experimental analysis of behavior*, *105*(1), 23–40.

Niv, Y., & Chan, S. (2011). On the value of information and other rewards. *Nature neuroscience*, *14*(9), 1095.

Peirce, J. W. (2007). Psychopypsychophysics software in python. *Journal of neuroscience methods*, *162*(1), 8–13.

Prokasy, W. F. (1956). The acquisition of observing responses in the absence of differential external reinforcement. *Journal of Comparative and Physiological Psychology*, *49*(2), 131.

Roper, K. L., & Zentall, T. R. (1999). Observing behavior in pigeons: The effect of reinforcement probability and response cost using a symmetrical choice procedure. *Learning and Motivation*, *30*(3), 201–220.

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33), 12252–12257.

Silver, D., Sutton, R. S., & Müller, M. (2008). Sample-based learning and search with permanent and transient memories. In *Proceedings of the 25th international conference on machine learning* (pp. 968–975).

Spetch, M. L., Belke, T. W., Barnet, R. C., Dunn, R., & Pierce, W. D. (1990). Suboptimal choice in a percentage-reinforcement procedure: Effects of signal condition and terminal-link length. *Journal of the experimental analysis of behavior*, *53*(2), 219–234.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1) (No. 1). MIT press Cambridge.

Vasconcelos, M., Monteiro, T., & Kacelnik, A. (2015). Irrational choice and the value of information. *Scientific reports*, *5*.

Wyckoff, L. B. (1952). The role of observing responses in discrimination learning. part i. *Psychological review*, *59*(6), 431.